Project Data Management Plan Template for I*d*MRC Projects

The **Project Data Management Plan** (PDMP) records the particulars of the way that data collected and created during the project are managed, principally for use, re-use and re-purposing.

The project DMP for your project must be based on this template and located in a publicly accessible and searchable place. The default location is an anonymous log-in page of the research project wiki.

The file name for the DMP for your project must conform to the I*d*MRC file-naming convention, and should be identified using the 'dmp' document abbreviation. A versioning system must be in force; this is catered for in the IdMRC file name coding scheme

The Project Data Management Plan and the Project Data Record Manifest should be considered a pair, and must be co-located and reciprocally associated.

The PDMP should be 'read-only', editing rights being limited to nominated members of the originating research project such as the principal investigator and the project manager and the data manager.

Example entries are given in italics; these should be deleted/overwritten during completion of the document.

Summary of Research Activity

Project name

e.g. Long And Technical Textual Evaluation (LATTE)

Period of Project

e.g. October 2009 – March 2011

Funder(s)

e.g. EPSRC

Lead and partner organizations

e.g. University of Bath (lead), University of Cambridge, University of Leeds

Data access summary

e.g. Access to data in directory path/to/secrets is restricted to Project staff (i.e. [names]) and, for the purposes of long-term curation, data repository staff. For full details, consult the Confidentiality Agreement (see below).

Receiving repository

e.g. The data from this Research Activity will be deposited according to the IdMRC Projects DMP (see below).

or

The data from this Research Activity will be deposited with the UK Data Archive in accordance with ESRC funding requirements.

Related documentation

- RCUK Policy and Code of Conduct on the Governance of Good Research Conduct
- The University of Bath Good Practice Guide for Research
- Engineering Research Data Management Plan Specification
- IdMRC Projects Data Management Plan

e.g.

- *Project Proposal: [wiki link]*
- *Project Plan: [wiki link]*
- *Confidentiality agreement with [name]: [wiki link]*
- *Participant consent forms: [wiki link], [physical location/contact name/contact details]*
- *IPR Statement: [wiki link]*
- *UK Data Archive deposit requirements: [wiki link]*

The location of the Project Record Manifest is recorded separately below.

Data re-use

Could the Research Activity's data requirements be met in whole or in part by existing data? If not, briefly indicate how this is known. If so, identify the data that could be used and indicate any foreseen issues with accessing the data; explain why new data are being generated as well, if applicable.

e.g. Access to the British Orthographic Dataset [link] will be needed for this research. Terms of access to this dataset have already been established, details of which are in the Project Proposal. As this dataset is derived from newspaper text, additional data will need to be generated in order to perform the comparison with engineering reports.

or

Access to the British Orthographic Dataset [link] would greatly assist this research. This Project appears to fall within a category for which free access is granted, although it is unclear what the licensing arrangements are. As this dataset relates to publications pre-1950, additional data will need to be generated in order to extend coverage to subsequent decades.

or

Data from the British Orthographic Dataset [link] may have provided suitable data for this Project, but access is only available on commercial terms and the data may not be used for research purposes.

or

[Data catalogue] was consulted and no suitable data were found to exist.

Relating new data to existing data

Describe how the newly generated data relates to the wider landscape of existing data.

e.g. This project represents interdisciplinary research across engineering, text mining and linguistic analysis. Parallel research has already been done in chemistry [link to dataset] and meteorology [link to dataset]…

State the measures that will be/have been taken to ensure integrability between newly generated data and existing data.

e.g.

- *This project will do X to ensure data quality, as this was the standard employed by Y and Z projects. Full details are provided below.*
- *Provenance will be tracked using industry-standard RDF. Full details are provided in the Project Record Manifest.*
- *Access logs and monthly SHA5 checksum tests will be used to ensure the data is not tampered with or corrupted for the duration of the Project.*
- *DocBook will be used in preference to TEI due to wider software support. Full details of formats used are given below.*
- *Timestamps, tool versions and region marks will be recorded in the DocBook markup as this is conventional within text mining research.*

Future use of data

List any bodies or groups which might be interested in the data, and the foreseeable contemporary or future uses to which they might be put.

e.g.

- *Researchers in [discipline]/interested in [research area/topic] could use this data for…*
- *Engineers interested in knowledge management*
- *The corporal analysis could be used as the basis for further linguistic comparisons.*

State the measures that will be/have been taken to prepare the data for these bodies/groups/uses.

e.g.

- *The Data Case will be packaged in DDI format, as this is conventional in social science research. Full details are provided below.*
- *The corpus will be loaded onto a server running in a virtual machine along with the text mining software … so that the corpus can be mined without the full text being exposed. Full details are provided below.*
- *Timestamps, tool versions and region marks will be recorded in the DocBook markup as this is conventional within text mining research.*

Project Record Manifest

Location of Project Record Manifest

e.g. [link to page or attachment in Project wiki space]

If the PRM does not use the RAID association method, describe the procedure for updating the PRM.

e.g. Whenever a new record is created, the creator must record the filename and X-drive path in the PRM and…

Data generation and manipulation

Give a detailed account how the data will be/have been generated and manipulated, including the methods, technology, conventions, coding schemes, etc. that will be/have been used.

This information can be provided

- *as a prose description here;*
- *as a commentary (here) on a RAID diagram included in the Project Record Manifest;*
- *within the Project Record Manifest as a commentary on a RAID diagram, in which case cite here the section of the Project Record Manifest in which the commentary occurs;*
- *by reference to a journal/conference paper, in which case provide the full citation and a link to a local copy (on the Project wiki, on the X drive or in OPuS).*

Data organization

Describe how the data will be/have been organized.

Files are named according to the I*d*MRC file naming convention *[link]*, which see for an explanation of the metadata it encodes.

In-file metadata fields (e.g. author, title, and so on in Microsoft Office formats) will be filled out for all formats that support them.

Other matters to discuss:

- *whether and why data will be/have been kept in separate tables or combined together in a spreadsheet or database;*
- *directory conventions;*
- *how the Data Case will be assembled and packaged for deposit.*

Data quality

Describe the quality assurance procedures and standards that will be/were used. If any data quality issues were encountered, list them and describe what was done to resolve them.

e.g. Questionnaires will be reviewed by the Work Package leader before being used…

…a trial run found that the recording quality of existing equipment was not sufficient for the automated analysis tools, so new equipment was purchased that produced better quality recordings…

Data structures and formats

Specify the information, tools or resources that would be needed to manipulate the Data Records or make them human-readable, along with any special instructions.

Hardware environment actually used

e.g. IBM-compatible PCs with 32-bit x86 processors, MacBook Pros

Software environment actually used

e.g. Windows XP and Windows 7 environments; Ubuntu 10.04 Server running in a VirtualBox virtual machine – [instructions on running the virtual machine]; NX 3 CAD system

What other environments, tools and libraries might support the data? (To be completed once Data Records have been made)

e.g. VirtualBox 3.x is known to run on both 32- and 64-bit architectures on Windows XP+, Mac OS X (Intel only), Linux and OpenSolaris… Microsoft Office documents can be opened with OpenOffice.org

List all data formats used, citing format specification documents if available/known. Provide an explanation of why these particular formats have been selected for use.

e.g.

- *Microsoft Excel Workbook 2000-2003: ubiquitous.*
- *DocBook 5.x: widely used by text mining community (see above)*

Procedure for updating

e.g. Every two months, each Work Package leader will compare the Project Record Manifest with this section to ensure that all software tools and file formats have been added; if any are missing, the Work Package leader will assign a researcher to add the missing information…

Data semantics

What conventions (schemas, ontologies, etc.) will be used to allow interpretation of data, and why?

e.g. Dublin Core Metadata Terms will be used for general descriptive metadata, due to widely available support. Materials test data will use the CEN CWA 16200 schemas for compatibility with ISO 6892-1:2009. Occupations will be classified according to the 2010 Standard Occupational Classification to enable integration with…

What additional information would an interested reader need in order to understand the Data Records?

e.g. latte3dat100908ab.xls Sheet 1: units for columns 2-6 are in millimetres; rig set up and calibrations can be found in [link]…

Procedure for updating

e.g. Whenever a Data Record is created, the data creator will update this section with the information needed to understand the record…

History of this DMP

Changes

- *[Date] – [Name]*
    - *Change 1*
    - *Change 2*

OR, if the master copy of the DMP is a wiki page

Copies

e.g.

- *[Date] – [Name] released a copy of this DMP as latte0dmp111009md32.pdf*

Review of this DMP

Scheduled reviews

- *[Date] – [Name]*

Completed reviews

- *[Date] – [Name]*
    - *Outcome 1*
    - *Outcome 2*

Authors' contact details

Name

- *Email address*
- *Telephone number/extension*
- *Work address if not University of Bath*